

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

1c868 U.S. PTO
09/909901
07/23/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 7月26日

出 願 番 号

Application Number:

特願2000-225911

出 願 人

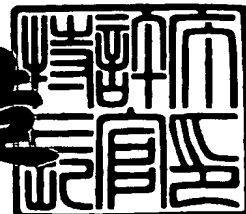
Applicant (s):

沖電気工業株式会社

2001年 3月30日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造



Inventors: Toshiki MURATA, et al

ANY DKT: 31759-173994 出証番号 出証特2001-3025137

【書類名】 特許願

【整理番号】 KN002336

【提出日】 平成12年 7月26日

【あて先】 特許庁長官 及川 耕造 殿

【国際特許分類】 G06F 17/27

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 北村 美穂子

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 村田 稔樹

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 佐々木 美樹

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 下畑 さより

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 福居 毅至

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 淵上 正睦

【特許出願人】

【識別番号】 000000295

【氏名又は名称】 沖電気工業株式会社

【代表者】 篠塚 勝正

【代理人】

【識別番号】 100090620

【弁理士】

【氏名又は名称】 工藤 宣幸

【手数料の表示】

【予納台帳番号】 013664

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9006358

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 自然言語処理装置及び方法

【特許請求の範囲】

【請求項 1】 少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理装置において、

予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる 1 以上の自然言語パターンを抽出する辞書引き手段と、

候補の自然言語パターンが木構造に適合しているか否かを検査するパターン検査手段と、

適合する場合に、その自然言語パターンを木構造に適用するパターン適用手段と

を有することを特徴とする自然言語処理装置。

【請求項 2】 予めパターン辞書に用意されている全て又は一部の上記自然言語パターンについて、パターン名及び又はパターン構成要素について、パターン適用条件を持たせておき、

上記パターン検査手段が、候補の自然言語パターンのパターン適用条件をも参照して木構造に適合しているか否かを検査する

ことを特徴とする請求項 1 に記載の自然言語処理装置。

【請求項 3】 パターン適用条件の 1 つとして意味条件に関する情報を有する複数の自然言語パターンのうちに、意味に関する情報のみが異なる複数の自然言語パターンを用意しておき、

上記パターン検査手段及び上記パターン適用手段の処理を通じて、意味情報が最適な木構造を決定する

ことを特徴とする請求項 2 に記載の自然言語処理装置。

【請求項 4】 予めパターン辞書に用意されている上記各自然言語パターンには、適用に対する優先度を表す情報が付与されていると共に、

木構造の適用の候補となっている上記自然言語パターンを、それに付与されている優先度の情報により評価するパターン評価手段をさらに有する

ことを特徴とする請求項 1 ～ 3 のいずれかに記載の自然言語処理装置。

【請求項 5】 パターン構成要素において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を有する自然言語パターンの優先度を高め、パターン名において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を持たない自然言語パターンの優先度を高めたことを特徴とする請求項 4 に記載の自然言語処理装置。

【請求項 6】 上記パターン評価手段は、木構造への適用に関し、パターン名及びそのパターン適用条件が同じであって優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、優先度が最も高い自然言語パターンの候補以外を候補から除くことを特徴とする請求項 4 又は 5 に記載の自然言語処理装置。

【請求項 7】 上記パターン評価手段は、木構造への適用に関し、パターン名及びパターン構成要素が同じであって優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、通常優先度より相対的に低い優先度の自然言語パターンの候補から除外することを特徴とする請求項 4 ～ 6 のいずれかに記載の自然言語処理装置。

【請求項 8】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、各木構造を評価し得る評価配点方法であって、複数の木構造で異なっている部分木に適用された自然言語パターンでの優先度の情報をも利用している評価配点方法に従って、複数の木構造の優劣を評価する木構造評価手段を有することを特徴とする請求項 4 ～ 7 のいずれかに記載の自然言語処理装置。

【請求項 9】 上記木構造評価手段は、木構造に適用された自然言語パターンでの優先度の情報に加え、複数の木構造で異なっている部分木を構成する終端記号の多少をも評価配点方法に反映させていることを特徴とする請求項 8 に記載の自然言語処理装置。

【請求項 10】 上記木構造評価手段は、木構造に適用された自然言語パターンでの優先度の情報に加え、所定の優先度に係るノードの位置をも評価配点方

法に反映させていることを特徴とする請求項 8 又は 9 に記載の自然言語処理装置。

【請求項 1 1】 自然言語パターンのユーザ登録手段を有することを特徴とする請求項 1 ～ 1 0 に自然言語処理装置。

【請求項 1 2】 ユーザ登録の自然言語パターンにシステム登録の自然言語パターン以上の優先度を付与することを特徴とする請求項 1 1 に記載の自然言語処理装置。

【請求項 1 3】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、複数の木構造で異なっている部分木において、ユーザ登録に係る自然言語パターンを適用された部分木を有する木構造を最優先する木構造評価手段を有することを特徴とする請求項 1 1 又は 1 2 に記載の自然言語処理装置。

【請求項 1 4】 少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理方法において、

予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる 1 以上の自然言語パターンを抽出する辞書引き工程と、

候補の自然言語パターンが木構造に適合しているか否かをチェックするパターン検査工程と、

適合する場合に、その自然言語パターンを木構造に適用するパターン適用工程と

を有することを特徴とする自然言語処理方法。

【請求項 1 5】 予めパターン辞書に用意されている全て又は一部の上記自然言語パターンについて、パターン名及び又はパターン構成要素について、パターン適用条件を持たせておき、

上記パターン検査工程が、候補の自然言語パターンのパターン適用条件をも参照して木構造に適合しているか否かをチェックする

ことを特徴とする請求項 1 4 に記載の自然言語処理方法。

【請求項 1 6】 パターン適用条件の 1 つとして意味条件に関する情報を有する複数の自然言語パターンのうちに、意味に関する情報のみが異なる複数の自然言語パターンを用意しておく、

上記パターン検査工程及び上記パターン適用工程の処理を通じて、意味情報が最適な木構造を決定する

ことを特徴とする請求項 1 5 に記載の自然言語処理方法。

【請求項 1 7】 予めパターン辞書に用意されている上記各自然言語パターンには、適用に対する優先度を表す情報が付与されていると共に、

木構造の適用の候補となっている上記自然言語パターンを、それに付与されている優先度の情報により評価するパターン評価工程をさらに有する

ことを特徴とする請求項 1 4 ～ 1 6 のいずれかに記載の自然言語処理方法。

【請求項 1 8】 パターン構成要素において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を有する自然言語パターンの優先度を高め、パターン名において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を持たない自然言語パターンの優先度を高めたことを特徴とする請求項 1 7 に記載の自然言語処理方法。

【請求項 1 9】 上記パターン評価工程は、木構造への適用に関し、パターン名及びそのパターン適用条件が同じであって優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、優先度が最も高い自然言語パターンの候補以外を候補から除くことを特徴とする請求項 1 7 又は 1 8 に記載の自然言語処理方法。

【請求項 2 0】 上記パターン評価工程は、木構造への適用に関し、パターン名及びパターン構成要素が同じであって優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、通常優先度より相対的に低い優先度の自然言語パターンの候補から除外することを特徴とする請求項 1 7 ～ 1 9 のいずれかに記載の自然言語処理方法。

【請求項 2 1】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、各木構造を評価し得る評価配点方法であって、複数の木構造で異なっている部分木に適用された自然言語パターンでの優先度の情報をも利

用している評価配点方法に従って、複数の木構造の優劣を評価する木構造評価工程を有することを特徴とする請求項 1 7 ～ 2 0 のいずれかに記載の自然言語処理方法。

【請求項 2 2】 上記木構造評価工程は、木構造に適用された自然言語パターンでの優先度の情報に加え、複数の木構造で異なっている部分木を構成する終端記号の多少をも評価配点方法に反映させていることを特徴とする請求項 2 1 に記載の自然言語処理方法。

【請求項 2 3】 上記木構造評価工程は、木構造に適用された自然言語パターンでの優先度の情報に加え、所定の優先度に係るノードの位置をも評価配点方法に反映させていることを特徴とする請求項 2 1 又は 2 2 に記載の自然言語処理方法。

【請求項 2 4】 自然言語パターンのユーザ登録工程を有することを特徴とする請求項 1 4 ～ 2 3 に自然言語処理方法。

【請求項 2 5】 ユーザ登録の自然言語パターンにシステム登録の自然言語パターン以上の優先度を付与することを特徴とする請求項 2 4 に記載の自然言語処理方法。

【請求項 2 6】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、複数の木構造で異なっている部分木において、ユーザ登録に係る自然言語パターンを適用された部分木を有する木構造を最優先する木構造評価工程を有することを特徴とする請求項 2 4 又は 2 5 に記載の自然言語処理方法。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は自然言語処理装置及び方法に関し、例えば、翻訳パターンを利用して機械翻訳を行う機械翻訳装置や機械翻訳方法などに適用し得るものである。

【 0 0 0 2 】

【従来の技術】

従来型の機械翻訳装置において、システムやユーザが登録可能な翻訳知識（シ

システム辞書やユーザ辞書) は、単語辞書、複数単語からなるイディオム、さらに、高性能なものでも「手を焼く」などの動詞と名詞の共起表現など、機械翻訳装置が予め決めている表現に限定されていた。これは、従来型の機械翻訳装置は、原言語の文法と辞書で原言語文を構文解析をし、変換辞書で言語変換し、目的言語の文法と辞書で目的言語文を生成するため、システムの文法に則った形式でしか、システムやユーザは翻訳辞書を登録できないためである。

【 0 0 0 3 】

さらに、原言語と目的言語は、個別に解析、生成されるため、翻訳知識としてユーザの直感に即した

” [名詞句] を次に示します。”

” F o l l o w i n g i s [名詞句] . ” (名詞句は変数) (P 1)
 のような原言語と目的言語の対から構成されるパターンに基づいた辞書も登録できない。

【 0 0 0 4 】

このような課題を解決したものとして、特開平 5 - 2 9 0 0 8 2 号公報に記載された翻訳パターンに基づく翻訳方法及び翻訳装置がある。この公報記載の発明は、翻訳知識を文脈自由文法の範囲内で表し、文法規則を原言語のパターンと目的言語のパターンとで対にしたものである。また、文法規則はトライ型の辞書に格納することにより、構文解析の速さを実用レベルにすることを可能にした。さらに、文法規則を原言語のパターンと目的言語のパターンとで対にしたことにより、同期導出が可能となり、言語変換、構文生成は簡単な処理だけで済むようになった。

【 0 0 0 5 】

これにより、ユーザの直感に即した上述したパターン P 1 のような原言語と目的言語の対から構成される翻訳パターンに基づいた辞書が登録できるようになり、文脈自由文法の範囲内なら任意の翻訳知識を登録でき、翻訳処理できるようになった。

【 0 0 0 6 】

【発明が解決しようとする課題】

しかしながら、上記公報では、複数の解（例えば構文解析結果）から最適な解を選ぶことについては、最小の構成数（パターン数）の解を選ぶとのみ記述されている。この場合には、構成数がパラメータとなっているので、構成要素の素性や構成要素間の関係を考慮した訳し分けを行うことができない。また、構成数（パターン数）が優先度を規定するものになっているが、単純なパラメータのため、優先度が正しくない恐れがある。

【 0 0 0 7 】

また、上記公報では、パターンに優先順位を付けても良いと記載されている。しかし、その明確な方式は示されておらず、どのようにパターンに優先順位を付与し、最適な解をとらえるかは開示されていない。

【 0 0 0 8 】

本発明は、以上の点を考慮してなされたものであり、適切な自然言語処理結果を選択し得る自然言語処理装置及び方法を提供しようとしたものである。

【 0 0 0 9 】

【課題を解決するための手段】

かかる課題を解決するため、第 1 の本発明は、少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理装置において、（１）予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる 1 以上の自然言語パターンを抽出する辞書引き手段と、（２）候補の自然言語パターンが木構造に適合しているか否かを検査するパターン検査手段と、（３）適合する場合に、その自然言語パターンを木構造に適用するパターン適用手段とを有することを特徴とする。

【 0 0 1 0 】

第 2 の本発明は、少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理方法において、（１）予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる 1 以上の自然言語パターンを抽出する辞書引き工程と、（２）候補の自然言語パターンが木構造に適合

しているか否かを検査するパターン検査工程と、（３）適合する場合に、その自然言語パターンを木構造に適用するパターン適用工程とを有することを特徴とする。

【 0 0 1 1 】

【発明の実施の形態】

（Ａ）第１の実施形態

以下、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第１の実施形態を図面を参照しながら詳述する。

【 0 0 1 2 】

（Ａ－１）第１の実施形態の構成

図１は、第１の実施形態の機械翻訳装置の機能的構成を示すブロック図である。なお、実際上は、例えば、パソコンなどの情報処理装置上に、処理プログラムや固定データなどがローディングされて、第１の実施形態の機械翻訳装置が構築される。

【 0 0 1 3 】

図１において、第１の実施形態の機械翻訳装置は、大きくは、入出力部１、翻訳処理部２及び辞書格納部３から構成される。

【 0 0 1 4 】

入出力部１は、キーボードやファイル読込装置等の入力部０１から翻訳対象となる入力文（原言語文）を受け付ける入力処理部１１と、翻訳した結果（目的言語文）をディスプレイやプリンタやファイル格納装置等の出力部０２に出力する出力処理部１２を有する。

【 0 0 1 5 】

翻訳処理部２は、入力文の形態素解析を行う形態素解析部２１、翻訳パターンを利用して原言語の構文解析を行う構文解析部２２、原言語の構文解析の結果に基づいて目的言語による構文の生成を行う構文生成部２３、及び、構文生成の結果に基づいて目的言語の形態素生成を行う形態素生成部２４からなる。

【 0 0 1 6 】

構文解析部２２は、パターン辞書引きを行う辞書引き部２２１と、辞書引きさ

れたパターン（原言語パターン）を構築中の木構造に適合するかを検査するパターン検査部 2 2 3 と、その検査結果によってパターンを構築中の木構造に適用するパターン適用部 2 2 4 と、これらの処理部を呼び出し木構造を構築する構文解析器 2 2 2 とを有する。

【 0 0 1 7 】

辞書格納部 3 は、原言語の形態素情報が格納されている原言語形態素辞書 3 1、目的言語の形態素情報が格納されている目的言語形態素辞書 3 2、及び、原言語と目的言語のパターンの対からなる翻訳パターンが格納されている翻訳パターン辞書（システムパターン辞書） 3 3 からなる。

【 0 0 1 8 】

（A-2）第 1 の実施形態の動作

次に、第 1 の実施形態の機械翻訳装置の翻訳動作（第 1 の実施形態の機械翻訳方法）を説明する。ここで、図 2 が、第 1 の実施形態の機械翻訳装置の翻訳動作を示すフローチャートである。

【 0 0 1 9 】

なお、第 1 の実施形態の機械翻訳装置は 1 文を入力単位としている。また以下の説明においては、“The policeman arrests him.” が入力された場合を例に動作を説明する。

【 0 0 2 0 】

ユーザが、上記入力文を入力部 0 1 により入力すると、入力処理部 1 1 が受け取り、翻訳処理部 2 の形態素解析部 2 1 に渡す（ステップ S 2 1）。形態素解析部 2 1 は、原言語形態素辞書 3 1 の情報を用いて、形態素解析処理を行う（ステップ S 2 2）。

【 0 0 2 1 】

形態素解析結果は、図 3 に示すように、ルートノードを“Node”とした木構造で表現される。複数候補がない形態素の場合には、ルートノードの直下に各形態素の標準形 3-1 と品詞や変化形などの形態素情報 3-2 とが付与される。一方、複数候補がある形態素の場合には、or ノード 3-3 の子ノードとして各形態素候補の情報が付与される。例えば、“arrests”は、動詞及び名詞

の品詞を取り得るので、複数候補を有する形態素であり、図 3 に示すように、各形態素候補の情報が付与される。

【 0 0 2 2 】

このような形態素解析結果が与えられた構文解析部 2 2 は、この結果に対し、翻訳パターン辞書 3 3 を用いて、構文解析処理を行う（ステップ S 2 3）。

【 0 0 2 3 】

図 4 は、翻訳パターン辞書 3 3 の格納内容（翻訳パターン）の一部を取り出して示したものである。

【 0 0 2 4 】

図 4 において、左側が英語パターンを表し、右側が日本語パターンを表し、これら英語パターン及び日本語パターンが対になっている。各言語のパターンは、

〔言語名：パターン名 パターン構成要素〕

からなる。

【 0 0 2 5 】

言語名は、英語（e n）か日本語（j a）かを規定する。

【 0 0 2 6 】

言語名に続くパターン名は、例えば、V P（動詞句）、N P（名詞句）、N（名詞）等の句構造規則での標識が適用される。

【 0 0 2 7 】

パターン構成要素は、単語、変数、又は、単語と変数の 2 以上の並び、からなる。変数は〔任意の数字：パターン名（木構造の下位ノードに対応する）〕で記述される。任意の数字部分は、対となっている原言語及び目的言語パターン間での対応関係を示すものである。構文解析においては、変数に、別のパターンが適用されることにより、パターンは入れ子構造をとることができる（変数が解消される）。また、単語及びパターン名は、符号 4 - 1 a や 4 - 1 b が付された部分のように、意味情報などの詳細な情報（素性情報）を持つことができる。さらに、単語及びパターン名は、符号 4 - 7 a が付された部分のように詳細情報を変数化して、情報の参照をすることもできる。

【 0 0 2 8 】

このような翻訳パターン（原言語パターン）を利用する第1の実施形態の構文解析処理は、図5のフローチャートに示すように、主として、3つの処理から構成され、ボトムアップに木構造を構築していくものである。

【0029】

すなわち、構文解析が終了していないことを確認しつつ（ステップS51）、パターン辞書引き処理（ステップS52）、パターン検査処理（ステップS53）及びパターン適用処理（ステップS54）の3つの処理を繰り返し行う。

【0030】

第1の処理であるパターン辞書引き処理（ステップS52）は、形態素解析結果及びそれまでのパターン適用処理の結果から、次に適用の可能性のある翻訳パターンを翻訳パターン辞書33から引く処理である。例えば、図3の”him:品詞=p r n”（符号3-4）という形態素解析結果より、図4の符号4-4を付した翻訳パターン（原言語パターン）が翻訳パターン辞書33から引かれる。

【0031】

第2の処理であるパターン検査処理（ステップS53）は、辞書引きの結果得られた翻訳パターンが現在構築中の木構造に適合するか否かを、各木構造毎に検査する処理である。

【0032】

第3の処理であるパターン適用処理（ステップS54）は、検査の結果、適合すると判定された木構造と翻訳パターンとに基づいて、木構造にそのパターンを実際に適用する処理である。

【0033】

図6及び図7を用いて、上述のパターン検査処理及びパターン適用処理を具体的に説明する。

【0034】

現在構築中の木構造が図6（a）であり、検査したい翻訳パターンが図6（c）であったとする。変数のパターン名は、NP同士で同じであり、NPがもつ詳細な情報も「意味=人」で一致する。従って、図6（c）に示す翻訳パターンは、図6（a）の木構造に適合するとパターン検査処理で判断され、パターン適用

処理によって、その翻訳パターンが適用され、図 7 に示す木構造が構築される。

【0035】

一方、現在構築中の木構造が図 6 (b) であり、検査したい翻訳パターンが図 6 (c) であった場合は、変数のパターン名は、NP 同士で同じであっても、NP の詳細情報（意味の情報）が異なるので、パターン検査処理で当該翻訳パターンは木構造に適合しないと判断され、当然に、パターン適用処理が実行されることもない。なお、図 6 (b) において、「意味！＝人」は「意味が人ではない」を表している。すなわち、「！」は否定を表している。

【0036】

入力文” The policeman arrests him. ” に対し、図 4 に示すような翻訳パターンを適用した構文解析結果は、図 8 に示すようになる。なお、この入力文に対しては、図 4 の翻訳パターン 4-1、4-3～4-7 が適用される。

【0037】

このような構文解析結果が構文生成部 23 に与えられ、構文生成部 23 によって、構文生成処理（図 2 のステップ S 24）が実行される。この構文生成処理では、翻訳パターン辞書 33 を参照し、原言語（英語）のパターンと対をなしている目的言語（日本語）のパターンを利用して、構文解析結果に対応する日本語の木構造を得る。図 9 は、上述した図 8 の構文解析結果に対応した構文生成結果（木構造）を示している。

【0038】

なお、翻訳パターンは、原言語パターンと翻訳パターンとの対になっており、その対応が一意であるので、実際上は、構文解析処理と構文生成処理とがほぼ並行して実行される。また、構文解析結果に対し、パターンを用いない従前からの方法によって構文生成を行っても良い。

【0039】

構文生成部 23 が生成した木構造は形態素生成部 24 に与えられ、形態素生成部 24 は形態素生成処理を行う（ステップ S 25）。形態素生成処理では、構文生成結果を、終端ノードに位置する日本語単語を左から順に並べ、各単語を目的

言語形態素辞書 32 を用いて、動詞の活用形を整えたりなど、各単語の整形を行う。その形態素生成結果（訳文）を出力処理部 12 が受取り、出力部 02 によって表示出力したり、ファイルに格納したりする（ステップ S26）。

【0040】

上述した入力文” The policeman arrests him. ”
に対しては、「警察官は彼を逮捕する。」という翻訳結果が得られる。

【0041】

ここで、上記入力文とは目的語だけが異なる” The policeman
arrests the deterioration. ” が入力された場合を
考える。

【0042】

この場合には、目的語にくる名詞の意味が人以外であるので、「警察官は悪化
を阻止する。」という翻訳結果が得られる。

【0043】

なお、この入力文に対する構文解析では、図 4 の翻訳パターン 4-2、4-3
、4-5～4-8 が適用される。

【0044】

すなわち、上述した図 6（a）及び図 6（b）に対する構築中の木構造に対し
、パターン 4-8、4-6 が入れ子に挿入された目的語（NP）が人以外を意味
するパターン 4-2 との適合判断により、図 6（b）が適合すると判定される。
その結果、パターン 4-2 の情報が盛り込まれた木構造が構文解析で得られ、こ
れにより、arrest の訳としてを阻止するが採用される。

【0045】

以上のように、「意味＝人」や「意味！＝人」などのパターン適用条件を、パ
ターンに盛り込んでおくことにより、同じ動詞などを使用した入力文に対して異
なる訳語を与えることができる。

【0046】

（A-3）第 1 の実施形態の効果

第 1 の実施形態の機械翻訳装置及び方法によれば、以下の効果を奏することが

できる。

【0047】

(a) 変数による入れ子構造の翻訳パターンを作ることができるだけでなく、変数や単語に意味などのパターン適用条件を与えることによって、条件による訳し分けをすることができる。

【0048】

従来技術の項で言及した公報では、“claims \$ and \$”、“claim \$”のように、条件に相当する情報を英語の表現レベルまで落していたが、この実施形態の場合には、英語の表現レベルまで落す必要はない。この実施形態では、“claims [NP: 数情報=複数]”、“claim [NP: 数情報=単数]”となる。

【0049】

(b) =や!=などパターン適用条件の記述が簡単であり、容易にパターンを作成することができる。例えば、目的語の意味が「人」ならば、訳はA、そうではないならば、訳はBというように、記述形式が人間の直観に即した形式であり、翻訳パターンの作成が容易である。

【0050】

(c) “I give these books.” は、構文的には、「私がこれらに本を与える。」と「私がこれらの本を与える。」の2通り存在するが、[en: VP give [1: NP: 意味=人] [2: NP]] という条件をもつパターンを登録することによって、前者の候補はあり得なくなる。このように、意味によるパターン適用条件をパターンに記述することによって、構文の曖昧性を削減することもできる。

【0051】

(A-4) 第1の実施形態の変形実施形態

上記では、構文解析結果に曖昧性が存在しない例文を用いて説明した。しかし、曖昧性が存在する場合（構文解析結果にorノードが生じる場合（第2の実施形態参照））にも、第1の実施形態を適用できる。例えば、orノードの子ノードのいずれか1つを選んで曖昧性のない構文解析結果を作ってそれを生成したり

、又は、曖昧性をもったまま構文生成を行い、複数の翻訳結果を出力したりしても良い。

【 0 0 5 2 】

第 1 の実施形態の技術思想は、特定の言語に依存せず、どんな言語間の翻訳にも適用可能である。

【 0 0 5 3 】

上記では、目的格に位置する名詞の意味による動詞の訳し分けの例を示したが、逆に動詞が持っている意味によって、名詞を訳し分ける場合にも適用できることは勿論であり、また、目的格だけでなく、主格による訳し分けも可能である。さらに修飾する語の意味や、前置詞句内の名詞句の意味による前置詞の訳し分け等、動詞や名詞以外の単語の訳し分けも可能である。さらに、名詞が単数形か複数形かによって名詞の訳語を決めるなど、意味以外のパターン適用条件による訳し分けも可能である。

【 0 0 5 4 】

パターン検査部などパターンに関する処理は、構文解析処理の外部に設けても良い（第 4 の実施形態参照）。

【 0 0 5 5 】

（B）第 2 の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第 2 の実施形態を図面を参照しながら詳述する。

【 0 0 5 6 】

（B-1）第 2 の実施形態の構成

図 1 0 は、第 2 の実施形態の機械翻訳装置の機能的構成を示すブロック図であり、上述した第 1 の実施形態に係る図 1 との同一、対応部分には同一符号を付して示している。

【 0 0 5 7 】

第 2 の実施形態の機械翻訳装置は、第 1 の実施形態の構成に加え、構文解析部 2 2 内にパターン評価部 2 2 5 を備えている。

【 0 0 5 8 】

パターン評価部 2 2 5 は、複数候補のパターンが存在する場合に、それらパターンを適用した複数の木構造を評価し、木構造を絞り込むものである。

【 0 0 5 9 】

なお、翻訳パターン辞書 3 3 に格納されている翻訳パターンも、第 1 の実施形態のものとは異なっている。この点については、動作の項で明らかにする。

【 0 0 6 0 】

(B - 2) 第 2 の実施形態の動作

この第 2 の実施形態においても、翻訳処理の全体の流れは、図 2 に示す通りであり、原文入力処理、形態素解析処理、構文解析処理、構文生成処理、形態素生成処理及び翻訳結果出力処理の順に実行される。

【 0 0 6 1 】

ここで、第 2 の実施形態は、第 1 の実施形態に比較して構文解析処理が異なっている。

【 0 0 6 2 】

図 1 1 は、第 2 の実施形態の機械翻訳装置での構文解析処理を示すフローチャートであり、第 1 の実施形態に係る図 5 との同一、対応処理ステップには同一符号を付して示している。

【 0 0 6 3 】

第 2 の実施形態の場合、パターン辞書引き処理（ステップ S 5 2）、パターン検査処理（ステップ S 5 3）、パターン適用処理（ステップ S 5 4）を順次行った後に、構築中の木構造ノードに対応して複数候補のパターンが存在するか否かを判別し（ステップ S 5 5）、複数候補のパターンが存在した場合に、パターン評価処理（ステップ S 5 6）を行なう。

【 0 0 6 4 】

新たに追加されたパターン評価処理の具体的な処理内容を説明する前に、図 1 2 を参照して、この第 2 の実施形態の翻訳パターン辞書 3 3 に格納されている翻訳パターンを説明する。

【 0 0 6 5 】

第 2 の実施形態の翻訳パターンは、先頭の” [” と末尾の”] ” と挟まれてい

る第1の実施形態と同様なパターンに加え、パターンの先頭に”！”が付与されたパターン12-9やパターンの先頭に”*”が付与されたパターン12-12がある。”！”や”*”は、パターンの優先度を表すための特別な記号である。

【0066】

図13は、パターンの優先度を規定するファイルの例であり、翻訳パターン辞書33又はパターン評価部22が格納しているものである。ここでは、ファイルの先頭から順に優先度が高いことを示している。また、[]は記号が何もついていない場合の順位である。従って、この例での優先順位は”！”が付与されているパターン、優先度の記号なしのパターン、”*”が付与されているパターンとなる。

【0067】

図12において、定冠詞の後に、名詞及び形容詞の両方を取り得る単語、さらに、名詞が続いている英語形態素列に関しては、2番目の単語を形容詞としたときには、パターン12-9が適用され、2番目の単語を名詞としたときには、パターン12-10が適用される。図12に示すように、パターン12-9がパターン12-10より優先度が上げられているので、動作の詳細は後述するが、両パターンが適用可能な場合であれば、パターン12-9が優先される。

【0068】

また、図12において、パターン12-11及びパターン12-12は共に文(S)に関するパターンであるが、意味条件が付されていない汎用的な方であるパターン12-12の優先度が低くなっている。動作の詳細は後述するが、意味条件が付されたパターン12-11が適用されない場合に、汎用的なパターン12-12が採用されることになる。

【0069】

パターン評価処理(ステップS56)では、このようなパターンの優先度を利用しながら複数候補のパターンが適用された複数の木構造の評価を行う。

【0070】

図14は、パターン評価処理の詳細を示すフローチャートである。パターン評価部224は、まず、ルートノードの詳細情報(パターン名の詳細情報)が一致

している複数の木構造候補があるか、ある場合には、ルートノードに高い優先度が付いた木構造候補があるか否かを判別する（ステップ S 1 4 1）。ここで、高い優先度とは候補の中で相対的に高い優先度を言い、必ずしも”！”が付与されたものに限定されない。そして、高い優先度が付いた木構造候補あれば、その木構造候補以外を候補から外す（ステップ S 1 4 2）。また、パターン評価部 2 2 4 は、子ノードの構成要素が一致している複数の木構造候補があるか、ある場合には、子ノードに低い優先度が付与されている木構造があるか否かを判別する（ステップ S 1 4 3）。ここで、低い優先度とは候補の中で相対的に低い優先度を言い、必ずしも”*”が付与されたものに限定されない。そして、低い優先度が付与されている木構造があれば、その木構造を候補から外す（ステップ S 1 4 4）。

【 0 0 7 1 】

次に、第 2 の実施形態の構文解析処理、特に、パターン評価処理について、入力文” He arrests the present deterioration. ”を例に説明する。この入力文に対する解析では、上述した図 1 2 に示した翻訳パターンが適宜利用される。

【 0 0 7 2 】

図 1 5 (a) 及び (b) はそれぞれ、入力文中の” the present deterioration ”の部分に対し、図 1 2 に示すパターン 1 2 - 9 及び 1 2 - 1 0 が適用されてできた候補（評価対象）としての木構造である。図 1 5 (a) に示す木構造には、翻訳パターン 1 2 - 9 に付与されている”！”の記号が付いている。パターン適用処理の終了後、図 1 5 (a) 及び (b) に示すような木構造のルートノードが等しく、かつ木構造を構成している単語が等しい木構造が存在すれば、これらの木構造がパターン評価処理に渡される。

【 0 0 7 3 】

パターン評価処理では、図 1 3 に示した優先度記号ファイルの内容を予め読み込んでおき、それに従って、パターンの適用を評価する。仮に、記号なしより高く評価したい記号がついた木構造が存在すれば、ルートノードがもつ詳細情報が等しいときのみ、最も高い記号をもつパターンのみを採用し、それ以外は構築途

中の木構造の候補から削除する。

【0074】

図15の例の場合には、図15(a)の木構造と図15(b)の木構造の詳細情報が等しいので、記号がついていない図15(b)の木構造は、優先度を高める記号が付いた図15(a)の木構造より優先度が低いので、木構造の候補から削除される。

【0075】

一方、図16は、記号なしのパターンより優先度を下げる記号”*”が付いている木構造を複数候補の中にもつ例である。優先度を下げる記号の評価の場合には、木構造の子ノードの構成要素が等しいことが条件となる。図16(a)に示すものはパターン12-11が適用された記号なし木構造であり、図16(b)に示すものはパターン12-12が適用された優先度を下げる記号”*”が付いた木構造であり、これら木構造は、子ノードの構成要素がNP、VPと等しく、かつ、図16(b)に示す木構造は”*”が付いているので、この木構造が候補から外される。

【0076】

図17は、入力文”He arrests the present deterioration.”に対する最終的な構文解析結果を示している。

【0077】

この構文解析結果に対応する日本語木構造が構文生成処理で生成され、形態素生成処理を経て、最終的に、「彼は現在の悪化を阻止する。」の翻訳結果が出力される。

【0078】

(B-3) 第2の実施形態の効果

第2の実施形態の機械翻訳装置及び方法によれば、第1の実施形態と同様な効果を奏すると共に、さらに、以下ような効果をも奏する。

【0079】

(a) 構文的には正しい複数の木構造候補が存在する場合、その中のある木構造を優先的に適用させて翻訳したい場合がある。この第2の実施形態では、優先

度を上げる記号”！”を優先的に適用させたいパターンに付与することで、これ
が実現できる。

【 0 0 8 0 】

(b) 木構造解析の途中で、パターンの条件は等しく、その優先度に差がある
複数候補のパターンが存在した場合に、優先度が低いパターンをその候補から外
すことにより、解析候補が減り、構文解析や構文生成にかかる処理時間を削減す
ることができる。

【 0 0 8 1 】

(c) 優先度下げ記号が付いた意味条件がついていないパターンを登録するこ
とによって、意味条件が適用する場合には、意味条件がありのパターンのみが適
用され、意味条件が適用しない場合にだけ、優先度下げ記号がついた意味条件な
しのパターンに適用するようになる。これにより、常に意味条件なしのパターン
を適用するという解析の無駄をなくすことができ、解析にかかる処理時間が削減
される。さらに、意味条件を適用しない場合も、構文解析の失敗にならず、尤も
らしい（汎用的な）訳語を得ることができる。

【 0 0 8 2 】

(d) 尤もらしい（汎用的な）訳語に対応したパターンに優先度を上げる記号
を付与することで、意味条件がないパターンに適用する場合には、厳しい条件を
もつ特殊な訳語ではなく、汎用的な訳語で翻訳することができる。

【 0 0 8 3 】

(B-4) 第2の実施形態の変形実施形態

優先度を示す記号はどんな種類の記号（単なる記号だけでなく、数字や括弧の
種類や文字列などを含む）であっても構わない。また、記号ではなく、他の形式
の情報であっても良い。

【 0 0 8 4 】

優先度を上げる記号として、！と＋など複数設け、！を優先させるというよう
に、優先度を順に複数用意することもできる。

【 0 0 8 5 】

(C) 第3の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第 3 の実施形態を図面を参照しながら詳述する。

【 0 0 8 6 】

(C-1) 第 3 の実施形態の構成

図 1 8 は、第 3 の実施形態の機械翻訳装置の機能的構成を示すブロック図であり、上述した第 2 の実施形態に係る図 1 0 との同一、対応部分には同一符号を付して示している。

【 0 0 8 7 】

第 3 の実施形態の機械翻訳装置は、第 2 の実施形態の構成に加え、木構造評価部 2 5 を備えている。

【 0 0 8 8 】

木構造評価部 2 5 は、構文解析により完成した原言語の構文解析結果として複数の候補が存在した場合に、どの候補が最も適切かをパターンがもつ情報などによって評価（点数計算）して、最適な候補（木構造）を選択して構文生成部 2 3 に与えるものである。又は、複数の候補を評価値順に並べ替えて構文生成部 2 3 に与えるものである。

【 0 0 8 9 】

(C-2) 第 3 の実施形態の動作

図 1 9 は、第 3 の実施形態の機械翻訳装置の翻訳動作を説明するフローチャートであり、第 1 の実施形態に係る図 5 との同一、対応処理ステップには同一符号を付して示している。

【 0 0 9 0 】

この第 3 の実施形態においては、原文入力処理（ステップ S 2 1）、形態素解析処理（ステップ S 2 2）、構文解析処理（ステップ S 2 3）、木構造評価処理（ステップ S 2 7）、構文生成処理（ステップ S 2 4）、形態素生成処理（ステップ S 2 5）及び翻訳結果出力処理（ステップ S 2 6）がこの順に実行される。すなわち、第 1 や第 2 の実施形態に比較して、木構造評価処理（ステップ S 2 7）が増えている。

【 0 0 9 1 】

図20は、木構造評価処理を示すフローチャートである。木構造評価部25は、構文解析結果を受け取ると、これをボトムアップに処理して評価点を計算する。木構造評価部25は、枝の先端側から非終端ノードを取り出しては（ステップS2001）、それがorノードかを判別し（ステップS2002）、orノードでなければ現ノードの点数を計算し（ステップS2005）、orノードであれば、子ノードの点数のうち最高点をノードの点数とした後（ステップS2003）、点数順にノードを並び換える（ステップS2004）。このような処理を、点数計算が終了していない非終端ノードがなくなるまで繰り返す。なお、具体的な点数の計算方法は後述する。

【0092】

以下、木構造評価処理を、入力文”It arrests an attention.”を例に具体的に説明する。

【0093】

図21は、この入力文に対する構文解析結果である。この図21は、orノードが2個存在するので、4個の構文解析結果候補（4個の木構造をまとめて示している）を含んでいる。図21において、点線で引き出された数字は木構造評価処理において計算された値であり、構文解析結果では含まれていない。なお、構文解析処理で用いられた翻訳パターン辞書33は第2の実施形態のもの（図12）に準じているとする。但し、優先度上げ記号として“!”と“+”との2種類があり、前者が2段階の優先度上げを表し、後者が1段階の優先度上げを表しているとする。

【0094】

図22は、評価点の計算方法の定義ファイルを示している。評価点計算方法定義ファイルは、例えば、木構造評価部25に格納されている。

【0095】

評価点計算方法定義ファイルは、評価項目名フィールド221と、評価部分フィールド222と、評価点フィールド223とからなる。

【0096】

評価項目名フィールド221には、パターンに係る優先度の上げ下げの記号“

！”や“+”や“*”や“node”（一般的な優先度ノード）だけでなく、「木構造の終端記号の数」を意味するterminalや、「ユーザ登録に係るもの」を意味するuser（例えば形態素辞書やパターン辞書に対する登録）などの評価項目名が記述されている。

【0097】

評価部分フィールド222には、自ノード（例えばorノード）より下の枝の全てが評価部分であることを表すtotalや、自ノード（例えばorノード）の直下の子ノードが評価部分であることを表すchildなどの評価部分が記述されている。

【0098】

評価点フィールド223には、評価点が記述されている。ここで、評価部分がtotalであるものに対しては、該当する評価項目（例えば記号）の数（Nとする）に応じた計算式（例えば、 $1000 - 100N$ ）で評価点が記述されている。なお、 $1000 - 100N$ は、数学的には、 $1000 - 100 \times N$ を意味している。評価部分がchildであるものに対しては、固定の評価点が記述されている。

【0099】

図22に示す評価点計算方法定義ファイルは、ユーザ登録に係る事項を最も評価し、次に、終端記号数を評価し、その後は、パターンの優先度の順位を評価するような構成となっている。なお、最高優先度記号“！”については、orノード直下でのノードに付されているか否かをも大きな評価ウェイトとしている。

【0100】

上述した図20に示した木構造評価処理において、ステップS2005の点数計算は、評価点計算方法定義ファイルの規定内容に従ってなされる。

【0101】

図21により具体的な評価点の計算方法を説明する。

【0102】

ノード21-1の+Nは、“+”記号が付与されており、また、非終端記号数がattentionだけの1個であるので、図22の定義ファイルにおける2

25及び230のレコードが適用されて評価点は90000001点となり、一方、ノード21-2は、“+”記号が付与されておらず、また、非終端記号数がattentionだけの1個であるので、図22の定義ファイルにおける225のレコードが適用されて評価点は90000000点となる。その結果、次のorノード21-3では、高い方のノード21-1の評価点が採用される。

【0103】

また、ノード21-4のVPは“arrests attention”が一塊りであって終端記号数（終端ノード数）も子ノード数の1個であるので、ファイルのレコード225及び228が適用されて評価点は90000900点となり、一方、ノード21-5のVPは、終端記号数（終端ノード数）が3個、子ノード数が4個、“+”記号が枝の先端側に1個あるので、ファイルのレコード225、228及び2210が適用されて、評価点は7000601点となる。その結果、次のorノード21-6ではノード21-4の点数が採用される。

【0104】

以上のようにして、評価点の計算が終了すると、ノードの順番を点数順に並び変える。このようにして、並べ替えた後の木構造を図23に示す。

【0105】

構文生成処理は、この木構造に対応する日本語（目的言語）での木構造を生成する。仮に、英語の木構造にorノードがあれば、orノードの子ノードで最も先頭に位置する英語ノードに対応する日本語ノードのみが選ばれて、生成処理が行われる。その後、形態素生成処理が行われ、その結果、「それは興味を引く。」という翻訳結果が出力される。

【0106】

なお、仮に、ユーザが複数の翻訳結果候補を出力して欲しい場合には、orノードの子ノード全てに対応する日本語木構造を利用して構文生成、形態素生成処理を行う。この場合の翻訳結果は、「それは（興味を引く | 注意を阻止する）。」となる。この場合、前者の方がシステムは評価していることを示している。

【0107】

(C-3) 第3の実施形態の効果

第 3 の実施形態の機械翻訳装置及び方法によれば、第 1 及び第 2 の実施形態と同様な効果を奏すると共に、これに加え、以下の効果を奏することができる。

【0 1 0 8】

により以下の効果を有する。

【0 1 0 9】

(a) 木構造全体で、パターンに付与されている種々の優先度記号や、パターンの詳細情報を用いて評価することにより、よりきめ細かな訳し分けや構文解析との絡みによる訳し分けが可能となる。

【0 1 1 0】

(b) この第 3 の実施形態によれば、種々の優先度記号やパターン詳細情報の優先順位を決めておくことにより、優先順位に従った翻訳結果を期待でき、翻訳結果も予測しやすい。

【0 1 1 1】

(c) 終端パターンの数など、優先度記号以外の条件も、同じ計算式で同時に評価することができる。

【0 1 1 2】

(d) 例えば、“…t o d a y…”の解析結果において、名詞の候補しか存在しない場合は名詞として翻訳するが、名詞と副詞の両方の候補が存在した場合には、副詞として翻訳したい、というような場合、優先したい品詞（この場合副詞）のパターンに優先度を上げる記号を付与することで、多品詞の優先順位を制御することができる。

【0 1 1 3】

(C-4) 第 3 の実施形態の変形実施形態

第 3 の実施形態で用いる評価点の計算方法を制御する優先度記号は、第 2 の実施形態の優先度記号を併用してもかまわなく、また、異なるようにしても良い。前者の場合において、記号の付与によってパターン評価処理で候補を削減できるものは削減し、それ以外を点数計算で優先度を決めるという、1 つの記号で二段階の優先度処理が行われることになる。上記説明は、このような場合を意識している。

【 0 1 1 4 】

評価点の計算方法を規定する記号は、いくつ付与してもどんな種類であっても構わない。また、複数種類の記号を同時に付与することも可能である。

【 0 1 1 5 】

点数の計算式及び各評価項目の配点は、どんな式でもどんな配点でも構わない。また、簡単に変更し得るようにしても良い。木構造全体の記号の数の優先順位より、o r ノード直下のノードに付与された記号の優先順位を高くするなど、木構造のレベルを考慮に入れた点数計算をすることも可能である。

【 0 1 1 6 】

終端ノードの数だけでなく、他の木構造やパターンの特徴に従って優先度を定めることができる。例えば、第 1 の実施形態の目的語の名詞句の意味の数が少ないものほど優先するように、記号を付与すると、複数のパターンが条件に適用した場合でも、常に、条件の厳しいパターンが優先されて翻訳されるというように、翻訳結果の出力に一定の原則を与えることができる。

【 0 1 1 7 】

(D) 第 4 の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第 4 の実施形態を図面を参照しながら詳述する。

【 0 1 1 8 】

(D-1) 第 4 の実施形態の構成

図 2 4 は、第 4 の実施形態の機械翻訳装置の機能的構成を示すブロック図であり、第 3 の実施形態に係る図 1 8 との同一、対応部分には同一符号を付して示している。

【 0 1 1 9 】

第 4 の実施形態の機械翻訳装置は、パターン検査部 2 5 1、パターン適用部 2 5 2 及びパターン評価部 2 5 3 の部分を構文解析部 2 2 の外部に独立したパターン処理部 2 6 として備え、構文生成部 2 3 内の構文生成器 2 3 1 が利用し得るようにしている。

【 0 1 2 0 】

すなわち、パターンの検査、適用、評価を構文解析時だけでなく、構文生成時
も行うことができるようにしている。

【0121】

なお、翻訳パターンとして、同一の原言語パターンに対応する目的言語パター
ンが複数のこともある。また、図4では示していないが、目的言語パターンにも
パターン適用条件などを当然に付与することができる。この第4の実施形態は、
このようなことを前提としている。

【0122】

(D-2) 第4の実施形態の動作

第4の実施形態が既述した実施形態と異なる動作は、構文生成処理である。そ
こで、以下では、図25のフローチャートを参照しながら、第4の実施形態の構
文生成処理での特徴的な処理を説明する。

【0123】

従来方式では、英語（原言語）の木構造の構築とほぼ並行して日本語（目的
言語）の木構造も構築されるため（単純なパターンを利用した木構造の置き換え
）、特別な構文生成処理は行われず。そのため、日本語の木構造においてパタ
ーン適用条件のずれが生じていてもそれを検知することができない。

【0124】

この第4の実施形態は、英語の木構造の構築とほぼ並行して構築される日本語
の木構造において、複数候補ノードが発生した場合に、パターン検査処理、パタ
ーン適用処理、パターン評価処理を行うことによって、複数候補を解消しようと
したものである。英語の木構造に対して、例えば、日本語の木構造を従来の構文
生成処理（パターンを用いた単純な置き換え）で得た仮りの日本語（目的言語）
の木構造に対して、図25が実行される。

【0125】

まず、構文生成器231は、ルートノードからトップダウンに仮の木構造から
非終端記号を1個取り出し（ステップS2501）、それがo r ノードか否かを
判別する（ステップS2502）。

【0126】

o r ノードがあれば、o r ノードの子ノードのパターンと自ノードを含めた親の木構造をパターン検査部 2 2 3 に渡し、パターン検査処理及びパターン適用処理を実行させる（ステップ S 2 5 0 3）。パターン検査処理が終了したときには、その検査結果に応じて、パターン適用部 2 2 4 によるパターン適用処理が実行される（ステップ S 2 5 0 4）。その後、パターン適用処理により o r ノードが解消されたか否かを判別し（ステップ S 2 5 0 5）、解消されていないければ、パターン評価部 2 2 5 にパターン評価処理を実行させる（ステップ S 2 5 0 7）。

【 0 1 2 7 】

取り出した非終端記号が o r ノードでない場合には、そのノードをそのまま適用させる（ステップ S 2 5 0 6）。

【 0 1 2 8 】

このような処理を非終端記号が取り出せなくなるまで繰り返し、日本語（目的言語）の木構造を完成（生成）させる。

【 0 1 2 9 】

なお、パターン検査処理、パターン適用処理及びパターン評価処理は、既述した実施形態とほぼ同様な処理である。

【 0 1 3 0 】

以下、入力文 “the intellectual woman” を例に、第 4 の実施形態での構文生成処理を説明する。

【 0 1 3 1 】

図 2 6 （a）は、この入力文に対する構文解析結果（英語での木構造）を示し、図 2 6 （b）は、それに対応する日本語での木構造を示している。図 2 6 （b）の日本語木構造は英語木構造の対応結果から作られた仮りの木構造である。この仮の木構造に対して、上述した図 2 5 に示す処理が実行される。

【 0 1 3 2 】

図 2 6 （b）に示す木構造では o r ノードが含まれており、パターン検査処理やパターン適用処理が行われる。

【 0 1 3 3 】

ここで、図 2 7 （b 1）に示すような翻訳パターンが記述されている場合であ

って、パターン検査処理及びパターン適用処理に供する親の木構造が図 2 7 (a) に示すような場合であれば、パターン検査でパターンは適合するという結果が得られ、図 2 7 (b 1) の日本語パターンが適用される。その結果、o r ノードは解消されて次のノードに進む。

【 0 1 3 4 】

しかし、図 2 7 (b 2) に示すような翻訳パターンも、すなわち、日本語パターンが英語パターンに対応して 2 種類記述されている場合であって、両パターンが適用できる場合であれば、パターン評価処理が実行される。

【 0 1 3 5 】

2 種類の日本語パターンには優先度記号が付与されているので、ここで、1 個に絞り込まれる。又は、優先順位が付与される。

【 0 1 3 6 】

図 2 8 は、上記入力文に対して優先順位を付与した場合の構文生成処理結果を示している。この結果が形態素生成部に渡されて、「理知的な女性」又は「理知的な女性 | 知力を要する女性」の翻訳結果が得られる。

【 0 1 3 7 】

(D - 3) 第 4 の実施形態の効果

第 4 の実施形態の機械翻訳装置及び方法によれば、既述した実施形態の効果に加えて、以下の効果を奏することができる。

【 0 1 3 8 】

(a) 原言語に関する情報だけでなく、目的言語の情報によっても訳し分けをすることができる。

【 0 1 3 9 】

(b) 上記翻訳パターン辞書による双方向の翻訳を実施した場合、原言語、目的言語パターンの両方に優先度を付与することによって、それぞれの優先度を解析側、生成側で再び利用することができる。

【 0 1 4 0 】

(D - 4) 第 4 の実施形態の変形実施形態

上記では、パターン生成時に複数候補を減らすという、第 2 の実施形態に類似

の方式を説明したが、構文解析器は全ての候補を求め、その後、構文生成木の点数計算をするという第3の実施形態に類似した方式をとっても良い。また、両方の方式をとっても良い。

【0141】

原言語から目的言語の翻訳の逆（つまり目的言語から原言語への翻訳）を考えた場合、目的言語のパターンに付与した優先度は解析時に利用できる。その場合、原言語側のパターンに付与した優先度は、生成時に利用できる。

【0142】

(E) 第5の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第5の実施形態を図面を参照しながら詳述する。

【0143】

(E-1) 第5の実施形態の構成

図29は、第4の実施形態の機械翻訳装置の機能的構成を示すブロック図であり、第4の実施形態に係る図24との同一、対応部分には同一符号を付して示している。

【0144】

この第5の実施形態は、入出力部1に、ユーザの入力する翻訳パターンを受け付けるユーザ辞書登録処理部13を備え、また、辞書格納部3にその翻訳パターンを格納するユーザパターン辞書34を備えている。すなわち、ユーザ登録（削除）機能を搭載したものである。

【0145】

そのため、構文解析部22などは、システムパターン辞書33とユーザパターン辞書34の両方を参照することになる。

【0146】

(E-2) 第5の実施形態の動作

ユーザ登録に係るものが翻訳パターンであるため、条件等の詳細な情報の直接的な登録は専門知識が必要とされるが、ユーザインターフェース等を備えることにより、単語（形態素）その他の要素の登録と同様な処理により登録することが

できる。但し、以下のような点が異なっている。

【0147】

”Following is [名詞句]”を“[名詞句]を次に示します。”をユーザが登録する場合について説明する。

【0148】

この場合の翻訳パターンは、

[en:S [following is [1:NP].]

[ja:S [1:NP]を次に示します。] となる。

【0149】

仮に、名詞句に位置する単語が人以外の場合のみこのパターンを適用したい場合には、[en:S [following is [1:NP:意味!=人].] [ja:S [1:NP]を次に示します。] というように意味条件も記述して登録する。この場合において、ユーザによる任意の入力を待ち受けるだけでなく、ユーザ辞書登録処理部13が、例えば、NPなどの場合に条件設定を行うか否かを問うメッセージや、条件の設定例などを表示して、条件を取り込むようにしても良い。

【0150】

また、ユーザ登録に係るパターンには、最高の優先度上げ記号を常に付加して登録することにする。

【0151】

さらに、上述したように、図22の評価点計算方法定義ファイル（第3の実施形態に係る）では、userという評価項目名のレコード221に最も高い評価点を付与するようにしており、木構造の評価時にユーザパターン辞書のパターンに係るものが最優先され、その翻訳結果が得られる。

【0152】

ここで、userはユーザパターンを示し、childは子ノードの直下を示す。レコード22-1は、orノード直下のユーザパターンに対し、最高点（100000000点）を与える、を意味する。これにより、ユーザパターンは、全ての候補に対して、優先度が高くなり、ユーザが登録したパターンは、必ず翻

訳結果に反映されるようになる。

【 0 1 5 3 】

(E-3) 第5の実施形態の効果

第5の実施形態の機械翻訳装置及び方法によれば、既述した実施形態の効果に加えて、以下の効果を奏することができる。

【 0 1 5 4 】

(a) ユーザが作成する翻訳パターン辞書にも、変数に意味やその他の条件を付与することができる。これにより、一般化したパターンによる登録ができ、登録するパターン数は少なくて済む。

【 0 1 5 5 】

(b) ユーザパターン毎に優先度記号を付与することも可能であり、ユーザによる訳語や訳文の制御も可能である。

【 0 1 5 6 】

(c) ユーザパターンに係る評価点も他のパターンに対する計算と同じ点数計算で行うため、ユーザパターン辞書の優先度（優先度記号）などの変更にも容易に対応することができる。

【 0 1 5 7 】

(d) ユーザは、候補訳語を全て出力することによって、ユーザパターンを用いない翻訳結果もユーザパターンを用いた翻訳と同時に得ることができる。

【 0 1 5 8 】

(E-4) 第5の実施形態の変形実施形態

第5の実施形態では、ユーザパターン辞書を最優先にしたが、最少ノード数を優先し、次に、ユーザパターンを優先する等、優先順位は、適宜変更することができる。

【 0 1 5 9 】

複数のユーザパターン辞書を作成して、それぞれの辞書に優先順位を付けることができる。

【 0 1 6 0 】

ユーザパターン辞書においても優先度を上げたり、下げたりする記号を付与す

ることができる。（その場合は、ユーザパターン辞書内における優先順位付けとなる。）

ユーザパターン辞書入力処理部において、ユーザは、[e n : …] [j a : …] などのような翻訳パターンを直接入力するのではなく、パターン名、単語と変数を簡単に入力するユーザインターフェースを設けることにより、ユーザパターン辞書の作成を容易にすることができる。

【0161】

第5の実施形態では、ユーザパターンの評価を木構造評価部で行うことによってユーザパターン辞書を優先して翻訳する方式を示したが、パターン評価部において、ユーザパターンとそれ以外の候補が存在した場合、ユーザパターンを適用して、それ以外の候補は適用しないという方式でも実現可能である（第2の実施形態参照）

（F）他の実施形態

本発明の技術思想の適用対象は、機械翻訳装置や方法に限定されず、自然言語パターンを利用して構文解析を行う自然言語処理装置及び方法や、自然言語パターンを利用して構文生成を行う自然言語処理装置及び方法に適用することができる。

【0162】

この場合において、構文解析を自然言語パターンを利用して行い、構文生成は自然言語パターンを利用しない方法で行うものであっても良く、その逆であっても良い。また、構文解析のみを必要とし、構文生成を実行しない装置や、その逆の装置などにも本発明を適用することができる。

【0163】

【発明の効果】

以上のように、本発明によれば、適切な自然言語処理結果を提供し得る自然言語処理装置及び方法を実現できる。

【図面の簡単な説明】

【図1】

第1の実施形態の機械翻訳装置の機能的構成を示すブロック図である。

【図 2】

第 1 の実施形態の機械翻訳装置の翻訳動作を示すフローチャートである。

【図 3】

第 1 の実施形態の形態素解析結果例を示す説明図である。

【図 4】

第 1 の実施形態の翻訳パターン辞書 3 3 の格納内容例を示す説明図である。

【図 5】

第 1 の実施形態の構文解析処理の詳細を示すフローチャートである。

【図 6】

第 1 の実施形態の構文解析処理のパターン検査処理及びパターン適用処理の具体例の説明図である。

【図 7】

図 6 のパターン適用処理後の木構造を示す説明図である。

【図 8】

第 1 の実施形態の入力文例に対する構文解析結果を示す説明図である。

【図 9】

図 8 の構文解析結果に対する構文生成結果を示す説明図である。

【図 1 0】

第 2 の実施形態の機械翻訳装置の機能的構成を示すブロック図である。

【図 1 1】

第 2 の実施形態の構文解析処理の詳細を示すフローチャートである。

【図 1 2】

第 2 の実施形態の翻訳パターン辞書 3 3 の格納内容例を示す説明図である。

【図 1 3】

第 2 の実施形態のパターンの優先度記号ファイルを示す説明図である。

【図 1 4】

第 2 の実施形態のパターン評価処理の詳細を示すフローチャートである。

【図 1 5】

第 2 の実施形態のパターン評価処理の対象となる木構造候補を示す説明図（1